

# Converse Bounds for Noisy Group Testing with Arbitrary Measurement Matrices

Jonathan Scarlett and Volkan Cevher

Laboratory for Information and Inference Systems (LIONS)

École Polytechnique Fédérale de Lausanne (EPFL)

Email: {jonathan.scarlett,volkan.cevher}@epfl.ch

**Abstract**—We consider the group testing problem, in which one seeks to identify a subset of defective items within a larger set of items based on a number of noisy tests. While matching achievability and converse bounds are known in several cases of interest for i.i.d. measurement matrices, less is known regarding converse bounds for arbitrary measurement matrices. We address this by presenting two converse bounds for arbitrary matrices and general noise models. First, we provide a strong converse bound ( $\mathbb{P}^{\text{error}} \rightarrow 1$ ) that matches existing achievability bounds in several cases of interest. Second, we provide a weak converse bound ( $\mathbb{P}^{\text{error}} \not\rightarrow 0$ ) that matches existing achievability bounds in greater generality.

## I. INTRODUCTION

The group testing problem consists of determining a small subset of “defective” items within a larger set of items  $\{1, \dots, p\}$ . This problem has a history in areas such as medical testing and fault detection, and has regained significant attention with following new applications in areas such as communication protocols [1], pattern matching [2], and database systems [3], and new connections with compressive sensing [4], [5].

Let the items be labeled as  $\{1, \dots, p\}$ , and let  $S$  be the subset of defective items. We consider a general group testing model in the observation  $Y$  associated with a single test is randomly generated according to

$$\mathbb{P}[Y = y \mid X = x, S = s] = P_{Y|X_S}(y|x_s) = P_{Y|V_S}(y|v_s), \quad (1)$$

where

$$V_S := \sum_{i \in S} X_i \quad (2)$$

counts the number of defective items in the test, and where the measurement vector  $X = (X_1, \dots, X_p) \in \{0, 1\}^p$  indicates which items are included in the test. While our techniques allow for arbitrary finite output alphabets, we focus on the binary case  $Y \in \{0, 1\}$  for concreteness. In the noiseless setting, we simply have  $Y = \mathbb{1}\{V_S > 0\}$ . Additive modulo-2 noise models of the form  $Y = \mathbb{1}\{V_S > 0\} \oplus Z$  are also common, but (1) is more general, permitting other forms of dependence on  $V_S$  such as that of dilution noise [6].

The goal is to recover  $S$  based on a number  $n$  of independent non-adaptive tests, with the  $i$ -th measurement vector being  $X^{(i)}$  and the  $i$ -th observation being  $Y^{(i)}$ . We henceforth let  $\mathbf{X}$  denote the  $n \times p$  matrix whose  $i$ -th row is  $X^{(i)}$ , and let  $\mathbf{Y}$  be the  $n$ -dimensional binary vector whose  $i$ -th entry is

$Y^{(i)}$ . We consider a fixed number  $k$  of defective items, and assume that the support set  $S$  is uniform over the subsets of  $\{1, \dots, p\}$  with cardinality  $k$ . For a fixed measurement matrix  $\mathbf{X}$ , the error probability is given by

$$P_e(\mathbf{X}) = \mathbb{P}[\hat{S} \neq S], \quad (3)$$

where  $\hat{S}$  is the estimate of  $S$  based on  $\mathbf{X}$  and  $\mathbf{Y}$ , and the probability is with respect to the randomness in  $S$  and  $\mathbf{Y}$ .

The information-theoretic limits of this problem have been studied for decades (e.g., see [7], [16]), and have recently become increasingly well-understood [8]–[13]. In particular, an exact asymptotic threshold is known in several cases of interest when we consider the error probability  $\bar{P}_e := \mathbb{E}[P_e(\mathbf{X})]$  averaged over an i.i.d. Bernoulli matrix  $\mathbf{X}$  with  $\mathbb{P}[X_{ij} = 1] = \nu/k$  ( $\nu > 0$ ). Specifically, in a broad range of scaling regimes with  $k = o(p)$ , we have  $\bar{P}_e \rightarrow 0$  if [12]

$$n \geq \max_{\ell=1, \dots, k} \frac{\ell \log \frac{p}{\ell}}{I(X_{s_{\text{dif}}}; Y|X_{s_{\text{eq}}})} (1 + \eta), \quad (4)$$

and  $\bar{P}_e \rightarrow 1$  if

$$n \leq \max_{\ell=1, \dots, k} \frac{\ell \log \frac{p}{\ell}}{I(X_{s_{\text{dif}}}; Y|X_{s_{\text{eq}}})} (1 - \eta). \quad (5)$$

In both of these equations,  $(s_{\text{dif}}, s_{\text{eq}})$  denotes an arbitrary partition of a fixed defective set  $s$  with  $|s_{\text{dif}}| = \ell$  (see [12] for further intuition), and the mutual information is with respect to the independent random vectors  $(X_{s_{\text{dif}}}, X_{s_{\text{eq}}})$  of sizes  $(\ell, k - \ell)$  containing independent Bernoulli( $\nu/k$ ) entries, and the model in (1) with  $s = s_{\text{dif}} \cup s_{\text{eq}}$ .

The main goal of this paper is to obtain variants of the converse bound with an additional optimization over  $\nu$ , i.e.,

$$n \leq \min_{\nu \in [0, k]} \max_{\ell=1, \dots, k} \frac{\ell \log \frac{p}{\ell}}{I(X_{s_{\text{dif}}}; Y|X_{s_{\text{eq}}})} (1 - \eta), \quad (6)$$

in the case of *arbitrary* measurement matrices, rather than i.i.d. measurement matrices.<sup>1</sup> We briefly mention some existing works in this direction:

- For the noiseless setting  $Y = \mathbb{1}\{V_S > 0\}$ , the threshold in (6) simplifies to  $(k \log_2 \frac{p}{k})(1 - \eta)$  [12], and the converse holds for arbitrary matrices by the so-called *counting bound* [14], [15].

<sup>1</sup> Although the measurement matrix  $\mathbf{X}$  may be arbitrary, our final results are still written in terms of random vectors  $X_{s_{\text{dif}}}$  and  $X_{s_{\text{eq}}}$  having independent Bernoulli entries. These are not *directly* related to  $\mathbf{X}$  itself.

- For the symmetric noise model  $Y = \mathbb{1}\{V_S > 0\} \oplus Z$  with  $Z \sim \text{Bernoulli}(\rho)$  for some  $\rho \in (0, 1)$ , the threshold in (5) simplifies to  $\frac{k \log_2 \frac{p}{k}}{\log 2 - H_2(\rho)}(1 - \eta)$  [12], where  $H_2(\rho) := -\rho \log \rho - (1 - \rho) \log(1 - \rho)$  is the binary entropy function in nats. Moreover, the converse remains valid for arbitrary matrices. This can be proved by combining the analysis of [12] with a simple symmetry argument on the information-density random variables, or can alternatively be obtained from a non-asymptotic bound given in [15].
- For general noise models, a *weak converse* statement corresponding to  $\ell = k$  (i.e.,  $s_{\text{eq}} = \emptyset$ ) is known for arbitrary matrices [7] (i.e., showing  $P_e(\mathbf{X}) \not\rightarrow 0$  as opposed to the strong converse  $P_e(\mathbf{X}) \rightarrow 1$ ).
- After the initial preparation of this work, we learned that a result similar to our second one (Theorem 2 below) was presented in the Russian literature [16, pp. 630-631], giving a weak converse for the case  $s_{\text{eq}} \neq \emptyset$ . However, the proof techniques appear to be significantly different, and the focus therein is on the case that  $k$  does not scale with  $p$ , in contrast with our work.

### A. Contributions

In this paper, we prove a strong converse corresponding to  $\ell = k$  for arbitrary matrices, and we prove a weak converse for all  $\ell = 1, \dots, k$ . Note that the former of these is of interest since  $\ell = k$  often achieves the maximum in (6); this is true for the noiseless model and the symmetric noise model [7], [12], and our numerical investigations suggest that it is also the case when  $P_{Y|V_S}$  corresponds to passing  $\mathbb{1}\{V_S > 0\}$  through a Z-channel [17]. However, there are known cases where only smaller values of  $\ell$  achieve the maximum [18].

### B. Notation

We write  $\mathbf{X}_S$  to denote the submatrix of  $\mathbf{X}$  containing the columns indexed by  $S$ . The complement with respect to the set  $\{1, \dots, p\}$  is denoted by  $(\cdot)^c$ , and similarly for  $X_S^{(i)}$ . For a given joint distribution  $P_{XY}$ , the corresponding marginal distributions are denoted by  $P_X$  and  $P_Y$ , and similarly for conditional marginals (e.g.,  $P_{Y|X}$ ). We use usual notations for the entropy and mutual information (e.g.  $H(X)$ ,  $I(X; Y|Z)$ ). We make use of the standard asymptotic notations  $O(\cdot)$ ,  $o(\cdot)$ ,  $\Theta(\cdot)$ ,  $\Omega(\cdot)$  and  $\omega(\cdot)$ . We define the function  $[\cdot]^+ = \max\{0, \cdot\}$ , and write the floor function as  $\lfloor \cdot \rfloor$ . The function  $\log$  has base  $e$ . The total variation (TV) distance between two probability mass functions is written as  $d_{\text{TV}}(P, Q)$ .

## II. STRONG CONVERSE FOR $s_{\text{eq}} = \emptyset$

Our first main result is as follows.

**Theorem 1.** *Consider any observation model  $P_{Y|V_S}$ , and define  $I_s^* := \max_{\nu \in [0, k]} I(X_s; Y)$ , where  $X_s$  has i.i.d.  $\text{Bernoulli}(\nu/k)$  entries. For any sequence of measurement matrices  $\mathbf{X}$  (indexed by  $p$ ), we have*

$$P_e(\mathbf{X}) \geq 1 - O\left(\frac{1}{n(I_s^*)^2}\right) \quad (7)$$

provided that

$$n \leq \frac{\log \binom{p}{k}}{I_s^*} (1 - \eta), \quad (8)$$

for arbitrarily small  $\eta > 0$ .

**Remark 1.** Typically in the case that  $s_{\text{eq}} = \emptyset$  we have  $I_s^* = \Theta(1)$ , and hence the remainder term  $O(\frac{1}{n(I_s^*)^2})$  behaves as  $O(\frac{1}{n})$ , in which case this lower bound on the error probability yields the strong converse statement  $P_e(\mathbf{X}) \rightarrow 1$ .

*Proof of Theorem 1:* Let  $\mathbf{X} \in \{0, 1\}^{n \times p}$  be a fixed measurement matrix. The analysis of [12] shows that

$$P_e(\mathbf{X}) \geq \sum_s \frac{1}{\binom{p}{k}} \mathbb{P} \left[ \sum_{i=1}^n \log \frac{P_{Y|X_S}(Y^{(i)}|X_S^{(i)})}{Q_Y(Y^{(i)})} \leq \log \binom{p}{k} + \log \delta_1 \mid \mathbf{X}, S = s \right] - \delta_1, \quad (9)$$

where  $Q_Y$  is an arbitrary auxiliary output distribution. Specifically, this was proved in [12] for the case that  $\mathbf{X}$  is i.i.d. and  $Q_Y$  is an induced output distribution, but the proof reveals this more general form.

Letting  $\mu_n(s)$  and  $\sigma_n^2(s)$  denote the mean and variance of  $\sum_{i=1}^n \log \frac{P_{Y|X_S}(Y^{(i)}|X_S^{(i)})}{Q_Y(Y^{(i)})}$  for a given defective set  $s$ , we obtain from Chebyshev's inequality that  $P_e \geq 1 - \sum_s \frac{1}{\binom{p}{k}} \frac{\sigma_n^2(s)}{(n\Delta I_s^*)^2} - \delta_1$  provided that  $\log \binom{p}{k} + \log \delta_1 \leq \mu_n(s) + n\Delta I_s^*$  for all  $s$ ; here  $\Delta \in (0, 1)$  is arbitrary for now.

The mean is directly computed as

$$\mu_n(s) = \sum_{i=1}^n \sum_y P_{Y|X_S}(y|X_S^{(i)}) \log \frac{P_{Y|X_S}(y|X_S^{(i)})}{Q_Y(y)} \quad (10)$$

$$= \sum_{i=1}^n \sum_y P_{Y|V_S}(y|V_S^{(i)}) \log \frac{P_{Y|V_S}(y|V_S^{(i)})}{Q_Y(y)} \quad (11)$$

$$= n \sum_{v_s, y} P_{V_S}^{(s)}(v_s) P_{Y|V_S}(y|v_s) \log \frac{P_{Y|V_S}(y|v_s)}{Q_Y(y)}. \quad (12)$$

where  $V_S^{(i)} := \sum_{j \in s} X_j^{(i)}$ , and  $P_{V_S}^{(s)}$  is the empirical distribution of  $V_S$  across the  $n$  tests for a given choice of  $s$ . Choosing  $Q_Y$  to be the unique capacity-achieving output distribution of the “channel”  $P_{Y|V_S}$ , it follows from (12) and a well-known saddlepoint result on the mutual information [19, Thm. 4.4] that, for all sets  $s$  having cardinality  $k$ , we have

$$\mu_n(s) \leq nI_s^*, \quad (13)$$

where  $I_s^*$  is defined in the theorem statement.

We claim that the corresponding variance behaves as

$$\sigma_n^2(s) = O(n). \quad (14)$$

This was shown for the case that  $Q_Y$  equals an induced output distribution in [11, App. A], but the analysis reveals that the same holds true for any  $Q_Y$  such that  $\min\{Q_Y(0), Q_Y(1)\}$  is bounded away from zero.

The proof of Theorem 1 is concluded by combining (13) and (14) with the above-mentioned application of Chebyshev's

inequality, and choosing  $\eta$  such that  $1 - \eta < \frac{1}{1+\Delta}$ , Since  $\Delta$  can be arbitrarily small, the same is true of  $\eta$ . ■

### III. WEAK CONVERSE FOR $s_{\text{eq}} \neq \emptyset$

Our second main result is as follows.

**Theorem 2.** *For any observation model  $P_{Y|V_S}$  and sequence of measurement matrices  $\mathbf{X}$  (indexed by  $p$ ), we have  $P_e(\mathbf{X}) \not\rightarrow 0$  provided that*

$$n \leq \max_{\ell=1, \dots, k} \min_{\nu \in [0, k]} \frac{\binom{p-k+\ell}{\ell}}{I(X_{s_{\text{dif}}}; Y|X_{s_{\text{eq}}}) + \Delta_\ell} (1 - \eta) \quad (15)$$

for some  $\eta > 0$ , where

$$\Delta_\ell = C_0 \frac{\ell(k-\ell)}{p} \max \left\{ 1, \log \frac{p}{\ell(k-\ell)} \right\} \quad (16)$$

for some universal constant  $C_0$ , and the mutual information is with respect to the pair  $(X_{s_{\text{dif}}}, X_{s_{\text{eq}}})$  having i.i.d. Bernoulli( $\nu/k$ ) entries, along with (1).

**Remark 2.** The remainder term  $\Delta_\ell$  is typically (but not always) dominated by the mutual information; for example, if the mutual information is  $\Theta(1)$  then this is true when  $k = O(p^\theta)$  for some  $\theta < \frac{1}{2}$ , regardless of the value of  $\ell$ .

**Remark 3.** The min-max ordering in (15) is the opposite of that in (6), thus making it a potentially weaker threshold. However, in the proof we also show that the threshold can be improved to

$$\min_{\mathcal{U}, P_U, P_{X|U}} \max_{\ell} \frac{\log \binom{p-k+\ell}{\ell}}{I(X_{s_{\text{dif}}}; Y|X_{s_{\text{eq}}}, U) + \Delta_\ell} (1 - \eta), \quad (17)$$

thus recovering the correct min-max ordering, but with an additional random variable  $U$  on a finite alphabet  $\mathcal{U}$ . This threshold can be shown to be achievable (hence establishing that (17) is a tight bound) in a broad range of scaling regimes using i-non-i.d. coding: Fix a sequence  $(u_1, \dots, u_n)$  with empirical distribution  $P_U$ , and then generate the  $i$ -th row according to an i.i.d. Bernoulli distribution  $P_{X|U}(\cdot|u_i)$  whose parameter  $\nu$  may depend on  $u_i$ . The achievability analysis then follows that in [12].

We have chosen to state the theorem in terms of the weakened threshold (15) since it bears a stronger resemblance to the more familiar threshold (6), and since we are not aware of any cases in which there is a gap between the two.

#### *Proof of Theorem 2:*

The proof is given in four steps.

*Step 1 (Fano's Inequality):* The starting point of our analysis is a necessary condition for  $P_e(\mathbf{X}) \rightarrow 0$  based on Fano's inequality and a genie argument, which follows directly from the analysis of [6] (see also [11, Sec. III-D]). Specifically, fixing  $\ell = 1, \dots, k$  and letting the *revealed indices* of  $S$  (denoted  $S_{\text{eq}}$ ) be uniform on the set of subsets  $\{1, \dots, p\}$  of size  $k - \ell$ , and letting the *non-revealed indices* of  $S$  (denoted

$S_{\text{dif}}$ ) be uniform on the set of subsets of  $\{1, \dots, p\} \setminus S_{\text{eq}}$  of size  $\ell$ , it is necessary that

$$1 \geq \frac{\log \binom{p-k+\ell}{\ell}}{I(S_{\text{dif}}; \mathbf{Y}|S_{\text{eq}})} (1 + o(1)). \quad (18)$$

We upper bound the mutual information by writing

$$I(S_{\text{dif}}; \mathbf{Y}|S_{\text{eq}}) \leq \sum_{i=1}^n I(S_{\text{dif}}; Y^{(i)}|S_{\text{eq}}) \quad (19)$$

$$= \sum_{i=1}^n I(V_{\text{dif}}^{(i)}; Y^{(i)}|V_{\text{eq}}^{(i)}), \quad (20)$$

where (19) is a standard property for independent observation models [17, Eq. (7.96)], and (20) follows by defining  $(V_{\text{dif}}^{(i)}, V_{\text{eq}}^{(i)})$  to count the number of defective items in the  $i$ -th test at the non-revealed and revealed indices, and recalling from (1) that  $Y$  depends on the defective set  $S = S_{\text{dif}} \cup S_{\text{eq}}$  only through  $V_S := \sum_{i \in S} X_i$ .

*Step 2 (Approximate Distributions by Binomials):* We proceed by showing that the pairs  $(V_{\text{dif}}^{(i)}, V_{\text{eq}}^{(i)})$  have a distribution which is “close enough” to a product of Binomial distributions with the same probability parameter. Since the defective set is uniformly random, the joint distribution of each pair  $(V_{\text{dif}}^{(i)}, V_{\text{eq}}^{(i)})$  (and hence the mutual information  $I(V_{\text{dif}}^{(i)}; Y^{(i)}|V_{\text{eq}}^{(i)})$ ) only depends on the number of non-zeros in the  $i$ -th row  $X^{(i)}$  of  $\mathbf{X}$ , which we denote by  $m^{(i)}$ .

Before proceeding, we recall that the Hypergeometric( $k, m, p$ ) distribution counts the number of “special items” obtained when sampling  $k$  items from a population of  $p$  items *without replacement*,  $m$  of which are labeled as special. A random variable with this distribution has probability mass function  $P_H(i) = \frac{\binom{m}{i} \binom{p-m}{k-i}}{\binom{p}{k}}$ . Of course, sampling *with replacement* simply gives the Binomial( $k, m/p$ ) distribution.

We have the following:

- 1) Recalling that  $S_{\text{eq}}$  is uniform on the  $\binom{p}{k-\ell}$  sets having cardinality  $k - \ell$ , the number of ones at the revealed indices is distributed as

$$V_{\text{eq}}^{(i)} \sim \text{Hypergeometric}(k - \ell, m^{(i)}, p). \quad (21)$$

We approximate this by the Binomial random variable

$$V_{\text{eq}}^{(i)} \sim \text{Binomial}\left(k - \ell, \frac{m^{(i)}}{p}\right). \quad (22)$$

Specifically, denoting the corresponding distributions by  $P_{V_{\text{eq}}}$  and  $P_{V_{\text{eq}}'}$  respectively (omitting the superscripts  $(\cdot)^{(i)}$ ), the total variation distance between the two satisfies [20]

$$d_{\text{TV}}(P_{V_{\text{eq}}}, P_{V_{\text{eq}}'}) \leq \frac{k - \ell - 1}{p - 1} = O\left(\frac{k - \ell}{p}\right). \quad (23)$$

We denote this upper bound by  $\delta_1$ .

- 2) Suppose that we condition on some value  $v_{\text{eq}}^{(i)}$  of  $V_{\text{eq}}^{(i)}$ . Recalling that  $(S_{\text{dif}}|S_{\text{eq}} = s_{\text{eq}})$  is uniform on the

$\binom{p-k+\ell}{\ell}$  possible realizations, we have

$$(V_{\text{dif}}^{(i)} | V_{\text{eq}}^{(i)} = v_{\text{eq}}) \sim \text{Hypergeometric}(\ell, m^{(i)} - v_{\text{eq}}, p - k + \ell). \quad (24)$$

We approximate this by the conditional distribution

$$(V_{\text{dif}}^{\dagger(i)} | V_{\text{eq}}^{(i)} = v_{\text{eq}}) \sim \text{Binomial}\left(\ell, \frac{m^{(i)} - v_{\text{eq}}}{p - k + \ell}\right), \quad (25)$$

which we further approximate by the unconditional distribution

$$V_{\text{dif}}^{\prime(i)} \sim \text{Binomial}\left(\ell, \frac{m^{(i)}}{p}\right). \quad (26)$$

Specifically, the corresponding distributions satisfy [20]

$$\begin{aligned} d_{\text{TV}}(P_{V_{\text{dif}}}(\cdot | v_{\text{eq}}), P_{V_{\text{dif}}^{\dagger}}(\cdot | v_{\text{eq}})) \\ \leq \frac{\ell - 1}{p - k + \ell - 1} = O\left(\frac{\ell}{p}\right), \end{aligned} \quad (27)$$

and (proved in the Appendix)

$$d_{\text{TV}}(P_{V_{\text{dif}}^{\dagger}}(\cdot | v_{\text{eq}}), P_{V_{\text{dif}}^{\prime}}) = O\left(\frac{\ell(k - \ell)}{p}\right) \quad (28)$$

uniformly in  $m^{(i)}$  and  $v_{\text{eq}}$ . Denoting these bounds by  $\delta_{2,1}$  and  $\delta_{2,2}$ , we obtain from the triangle inequality that

$$d_{\text{TV}}(P_{V_{\text{dif}}}(\cdot | v_{\text{eq}}), P_{V_{\text{dif}}^{\prime}}) \leq \min\{1, \delta_{2,1} + \delta_{2,2}\} =: \delta_2, \quad (29)$$

where the upper bound of one is trivial.

#### Step 3 (Infer Bounds on the Mutual Informations)

Next, we formalize the statement that if two joint distributions are close in TV distance, their (conditional) mutual informations are also close. Using the above definitions of  $(V_{\text{dif}}, V_{\text{eq}})$ ,  $(V_{\text{dif}}^{\prime}, V_{\text{eq}}^{\prime})$  and  $(\delta_1, \delta_2)$ , we have the following:

- 1) We prove in the Appendix that

$$|I(V_{\text{dif}}; Y | V_{\text{eq}}) - I(V_{\text{dif}}; Y | V_{\text{eq}}^{\prime})| \leq \delta_1 \log 2. \quad (30)$$

- 2) We also prove in the Appendix that

$$|I(V_{\text{dif}}; Y | V_{\text{eq}}^{\prime}) - I(V_{\text{dif}}^{\prime}; Y | V_{\text{eq}}^{\prime})| \leq \delta_2 \log \frac{4}{\delta_2}. \quad (31)$$

In fact, we show that the logarithmic term can usually be improved to a constant and sometimes even  $o(1)$ ; see Remark 4. We focus on the slightly looser bound (31) for the sake of simplicity.

- 3) Combining these with (23), (27) and (28) gives

$$\begin{aligned} |I(V_{\text{dif}}; Y | V_{\text{eq}}) - I(V_{\text{dif}}^{\prime}; Y | V_{\text{eq}}^{\prime})| \\ = O\left(\frac{\ell(k - \ell)}{p} \max\left\{1, \log \frac{p}{\ell(k - \ell)}\right\}\right). \end{aligned} \quad (32)$$

Substituting (32) into (18) and (20), and maximizing over  $\ell$ , we obtain the necessary condition

$$n \geq \max_{\ell} \frac{\log \binom{p-k+\ell}{\ell}}{\frac{1}{n} \sum_{i=1}^n I(V_{\text{dif}}^{\prime(i)}; Y^{(i)} | V_{\text{eq}}^{\prime(i)}) + \Delta_{\ell}} (1 + o(1)) \quad (33)$$

where  $\Delta_{\ell} = O\left(\frac{\ell(k - \ell)}{p} \max\{1, \log \frac{p}{\ell(k - \ell)}\}\right)$ .

#### Step 4 (Form a Single-letter Expression)

By defining a random variable  $U$  equiprobable on  $\{1, \dots, n\}$ , we can write the average in the denominator of (33) as

$$\frac{1}{n} \sum_{i=1}^n I(V_{\text{dif}}^{\prime(i)}; Y^{(i)} | V_{\text{eq}}^{\prime(i)}) = I(V_{\text{dif}}^{\prime}; Y | V_{\text{eq}}^{\prime}, U), \quad (34)$$

where the conditional distributions of  $V_{\text{dif}}^{\prime}$  and  $V_{\text{eq}}^{\prime}$  given  $U = i$  are independent Binomial random variables with  $(\ell, k - \ell)$  trials and a common parameter  $\frac{m^{(i)}}{p}$ . Thus, the overall bound becomes

$$n \geq \max_{\ell} \frac{\log \binom{p-k+\ell}{\ell}}{I(V_{\text{dif}}^{\prime}; Y | V_{\text{eq}}^{\prime}, U) + \Delta_{\ell}} (1 + o(1)) \quad (35)$$

Upper bounding the right-hand side by maximizing over  $P_U$  and  $P_{X|U}$  yields (17); once again, since the output depends on the measurement vector  $X$  only through  $\sum_{i \in S} X_i$ , we can safely replace the Binomial random variables  $(V_{\text{dif}}^{\prime}, V_{\text{eq}}^{\prime})$  by the corresponding i.i.d. Bernoulli vectors  $(X_{s_{\text{eq}}}, X_{s_{\text{dif}}})$  in the mutual information. Further weakening (17) by swapping the min-max ordering yields (15), thus concluding the proof of Theorem 2. ■

## IV. CONCLUSION

We have provided two converse bounds for noisy group testing with arbitrary measurement matrices. Our first result strengthens an existing result [7] to obtain a strong converse statement  $P_e(\mathbf{X}) \rightarrow 1$ , and our second result provides a (weak) converse with a potentially improved threshold. In several cases, these converse bounds are known to be achievable using i.i.d. matrices when  $k$  scales sufficiently slowly compared to  $p$  [12], and thus our results support the use of such matrices in these regimes. In contrast, it is known that i.i.d. matrices can be suboptimal in other settings, such as the linear scaling  $k = \Theta(p)$  [13]. In such cases, there may be room to improve the converse bounds presented in this paper.

Another direction for future work is to determine to what extent our bounds remain valid in the case of adaptive group testing, where each test can be designed based on past observations. Some work in this direction is given in [15], but the most conclusive results therein are limited to symmetric noise.

## APPENDIX

### A. Proof of (28)

Recall that we are considering the TV distance between  $(V_{\text{dif}}^{\dagger} | V_{\text{eq}} = v_{\text{eq}}) \sim \text{Binomial}(\ell, \frac{m^{(i)} - v_{\text{eq}}}{p - k + \ell})$  and  $V_{\text{dif}}^{\prime} \sim \text{Binomial}(\ell, \frac{m^{(i)}}{p})$ . We define the difference between the two binomial parameters as

$$\Delta := \frac{m^{(i)}}{p} - \frac{m^{(i)} - v_{\text{eq}}}{p - k + \ell}. \quad (36)$$

By a simple asymptotic expansion and the fact that  $v_{\text{eq}} \in [0, k - \ell]$ , this satisfies

$$\Delta = O\left(\frac{k - \ell}{p}\right) \quad (37)$$



uniformly in  $m^{(i)}$  and  $v_{\text{eq}}$ . Moreover, the bound for comparing Binomial distributions in [21, Eq. (16)] states that

$$d_{\text{TV}}(P_{V_{\text{dif}}^\dagger}(\cdot|v_{\text{eq}}), P_{V_{\text{dif}}'}) \leq c\sqrt{\eta}(1 + \sqrt{2\eta})e^{2\eta}, \quad (38)$$

where  $c = (2\pi)^{1/4}e^{1/24}2^{-1/2}$  and  $\eta = \Delta^2\ell(\ell+2) = O(\Delta^2\ell^2)$ . This upper bound behaves as  $O(\sqrt{\eta}) = O(\frac{\ell(k-\ell)}{p})$  whenever  $\eta = O(1)$ , thus establishing (28). If  $\eta = \Omega(1)$ , then (28) is trivial anyway, since it gives  $\frac{\ell(k-\ell)}{p} = \Omega(1)$ , but an upper bound of 1 always holds.

### B. Proof of (30)

We obtain (30) by writing

$$|I(V_{\text{dif}}; Y|V_{\text{eq}}) - I(V_{\text{dif}}; Y|V_{\text{eq}}')| = \left| \sum_{v_{\text{eq}}} (P_{V_{\text{eq}}}(v_{\text{eq}}) - P_{V_{\text{eq}}'}(v_{\text{eq}})) I(V_{\text{dif}}; Y|v_{\text{eq}}) \right| \quad (39)$$

$$\leq \sum_{v_{\text{eq}}} |P_{V_{\text{eq}}}(v_{\text{eq}}) - P_{V_{\text{eq}}'}(v_{\text{eq}})| \log 2 \quad (40)$$

$$= d_{\text{TV}}(P_{V_{\text{eq}}}, P_{V_{\text{eq}}'}) \log 2, \quad (41)$$

where (40) holds since the mutual information is upper bounded by  $\log 2$  with binary outputs.

### C. Proof of (31)

Since the conditional mutual information is an average of unconditional mutual informations and (28) is uniform in  $v_{\text{eq}}$ , it suffices to show that for any  $P(x)$  and  $Q(x)$  on some common alphabet  $\mathbf{X}$ , the inequality  $d_{\text{TV}}(P, Q) \leq \delta$  implies  $|I_P(X; Y) - I_Q(X; Y)| \leq \delta \log \frac{4}{\delta}$ . Here the subscripts  $P$  and  $Q$  denote which distribution on  $X$  is used, whereas the conditional distribution  $W(y|x)$  of  $Y$  given  $X$  is the same in both cases. We use similar notations for entropies, such as  $H_P(Y)$  and  $H_P(Y|X)$ .

Since  $I(X; Y) = H(Y) - H(Y|X)$ , we have

$$|I_P(X; Y) - I_Q(X; Y)| \leq |H_P(Y) - H_Q(Y)| + |H_P(Y|X) - H_Q(Y|X)|. \quad (42)$$

For the second term, we follow (39)–(41) to deduce that

$$|H_P(Y|X) - H_Q(Y|X)| \leq d_{\text{TV}}(P, Q) \log 2. \quad (43)$$

Moreover, the same reasoning along with the identities  $P_Y(y) = \sum_x P_X(x)P_{Y|X}(y|x)$  and  $P_{Y|X}(y|x) \leq 1$  gives

$$d_{\text{TV}}(PW, QW) \leq d_{\text{TV}}(P, Q), \quad (44)$$

where  $PW$  denotes the  $Y$ -marginal of  $P(x)W(y|x)$ , and similarly for  $QW$ . We may thus apply the result on the continuity of entropy in [22, Ch. 2] to obtain

$$|H_P(Y) - H_Q(Y)| \leq d_{\text{TV}}(P, Q) \log \frac{2}{d_{\text{TV}}(P, Q)}. \quad (45)$$

Combining the above estimates yields  $|I_P(X; Y) - I_Q(X; Y)| \leq \delta \log \frac{4}{\delta}$  whenever  $d_{\text{TV}}(P, Q) \leq \delta$ , as desired.

**Remark 4.** The logarithmic factor in (45) can be replaced by a constant whenever  $P$  and  $Q$  yield probabilities of  $Y = 0$

and  $Y = 1$  that are strictly bounded away from one. This is because the entropy has bounded derivatives except as  $P_Y(y) \rightarrow 0$ . In fact, in the vicinity of  $P_Y = \{0.5, 0.5\}$  (which is relevant for symmetric settings), we may even make the bound in (45) behave as  $o(d_{\text{TV}}(P, Q))$ , since the derivative of the binary entropy function at 0.5 is zero.

### ACKNOWLEDGMENT

This work was supported by the European Commission (ERC Future Proof), SNF (200021-146750 and CRSII-2-147633), and ‘EPFL Fellows’ program (Horizon2020 665667).

### REFERENCES

- [1] A. Fernández Anta, M. A. Mosteiro, and J. Ramón Muñoz, “Unbounded contention resolution in multiple-access channels,” in *Distributed Computing*. Springer Berlin Heidelberg, 2011, vol. 6950, pp. 225–236.
- [2] R. Clifford, K. Efremenko, E. Porat, and A. Rothschild, “Pattern matching with don’t cares and few errors,” *J. Comp. Sys. Sci.*, vol. 76, no. 2, pp. 115–124, 2010.
- [3] G. Cormode and S. Muthukrishnan, “What’s hot and what’s not: Tracking most frequent items dynamically,” *ACM Trans. Database Sys.*, vol. 30, no. 1, pp. 249–278, March 2005.
- [4] A. Gilbert, M. Iwen, and M. Strauss, “Group testing and sparse signal recovery,” in *Asilomar Conf. Sig., Sys. and Comp.*, Oct. 2008, pp. 1059–1063.
- [5] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, “One sketch for all: Fast algorithms for compressed sensing,” in *Proc. ACM-SIAM Symp. Disc. Alg. (SODA)*, New York, 2007, pp. 237–246.
- [6] G. Atia and V. Saligrama, “Boolean compressed sensing and noisy group testing,” *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, March 2012.
- [7] M. Malyutov, “The separating property of random matrices,” *Math. notes Acad. Sci. USSR*, vol. 23, no. 1, pp. 84–91, 1978.
- [8] G. Atia and V. Saligrama, “A mutual information characterization for sparse signal processing,” in *Int. Colloq. Aut., Lang. and Prog. (ICALP)*, Zürich, 2011.
- [9] V. Tan and G. Atia, “Strong impossibility results for sparse signal processing,” *IEEE Sig. Proc. Letters*, vol. 21, no. 3, pp. 260–264, March 2014.
- [10] T. Laarhoven, “Asymptotics of fingerprinting and group testing: Tight bounds from channel capacities,” *IEEE Trans. Inf. Forens. Sec.*, vol. 10, no. 9, pp. 1967–1980, 2015.
- [11] J. Scarlett and V. Cevher, “Limits on support recovery with probabilistic models: An information-theoretic framework,” 2015, <http://infoscience.epfl.ch/record/204670>.
- [12] —, “Phase transitions in group testing,” in *Proc. ACM-SIAM Symp. Disc. Alg. (SODA)*, 2016.
- [13] M. Aldridge, “The capacity of Bernoulli nonadaptive group testing,” 2015, <http://arxiv.org/abs/1511.05201>.
- [14] L. Baldassini, O. Johnson, and M. Aldridge, “The capacity of adaptive group testing,” in *IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2676–2680.
- [15] O. Johnson, “Strong converses for group testing in the finite blocklength regime,” 2015, <http://arxiv.org/abs/1509.06188>.
- [16] M. B. Malyutov, “Search for sparse active inputs: A review,” in *Inf. Theory, Comb. and Search Theory*, 2013, pp. 609–647.
- [17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2001.
- [18] M. B. Malyutov and P. S. Mateev, “Screening designs for non-symmetric response function,” *Mat. Zametki*, vol. 29, pp. 109–127, 1980.
- [19] Y. Polyanskiy and Y. Wu, “Lecture notes on information theory,” 2014, [http://people.lids.mit.edu/yp/homepage/data/itlectures\\_v2.pdf](http://people.lids.mit.edu/yp/homepage/data/itlectures_v2.pdf).
- [20] S. Y. T. Soon, “Binomial approximation for dependent indicators,” *Statistica Sinica*, vol. 6, no. 3, pp. 703–714, 1996.
- [21] B. Roos, “Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion,” *Theory of Probability & Its Applications*, vol. 45, no. 2, pp. 258–272, 2001.
- [22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.